

ICMS

IJSN - Instituto Jones dos Santos Neves

2010

10

TEXTO PARA DISCUSSÃO

**MODELOS MENSAL E TRIMESTRAL PARA
PROJEÇÃO DE ARRECADAÇÃO DO ICMS
PARA O ESTADO DO ESPÍRITO SANTO**

GOVERNO DO ESTADO DO ESPÍRITO SANTO
SECRETARIA DE ESTADO DE ECONOMIA E PLANEJAMENTO - SEP
INSTITUTO JONES DOS SANTOS NEVES - IJSN

**TEXTO PARA
DISCUSSÃO** | **10**

MODELOS MENSAL E TRIMESTRAL PARA PROJEÇÃO DE ARRECADAÇÃO DO ICMS PARA O ESTADO DO ESPÍRITO SANTO

Livio Ribeiro
Doutorando em Economia
PUC-Rio

Vitória | 2010

Instituto Jones dos Santos Neves
Modelos mensal e trimestral para projeção
de arrecadação do ICMS para o estado do Espírito Santo.
Vitória, ES, 2009.

25p. il. (Texto para discussão, 10)
ISBN 978-85-62509-14-8

1.Economia. 2.ICMS. 3.Espírito Santo (Estado). I.Título.
II.Séries

Sumário

Apresentação	4
1. Introdução	5
2. Modelo Básico e Metodologia de Estimação	6
3. Resultados	10
3.1. Modelos AR (p)	10
3.2. Modelos OLS com Defasagem da Variável Dependente	13
3.3. Modelos DOLS	18
3.4. Escolha do Modelo de Projeção	21
4. Projeção e Conclusão	23
5. Referências Bibliográficas	25

APRESENTAÇÃO

O objetivo do presente estudo é fornecer uma metodologia inicial, com base em modelos econométricos, para projetar as receitas de ICMS do Estado do Espírito Santo. Foram comparados três tipos distintos de metodologia de estimação, modelos da classe ARIMA, modelos estruturais (que utilizam variáveis explicativas como produção industrial, vendas no varejo e importações) estimados por *Ordinary Least Squares* (OLS), com defasagens da própria receita de ICMS, e modelos estruturais estimados por *Dynamic Ordinary Least Squares* (DOLS). Todos os modelos foram avaliados em duas frequências de dados, mensais e trimestrais. Os resultados obtidos sugerem que os modelos trimestrais dominam estritamente os modelos mensais e que, dentre estes, os melhores modelos para projeção são os estruturais estimados por OLS com defasagem de um trimestre das receitas de ICMS.

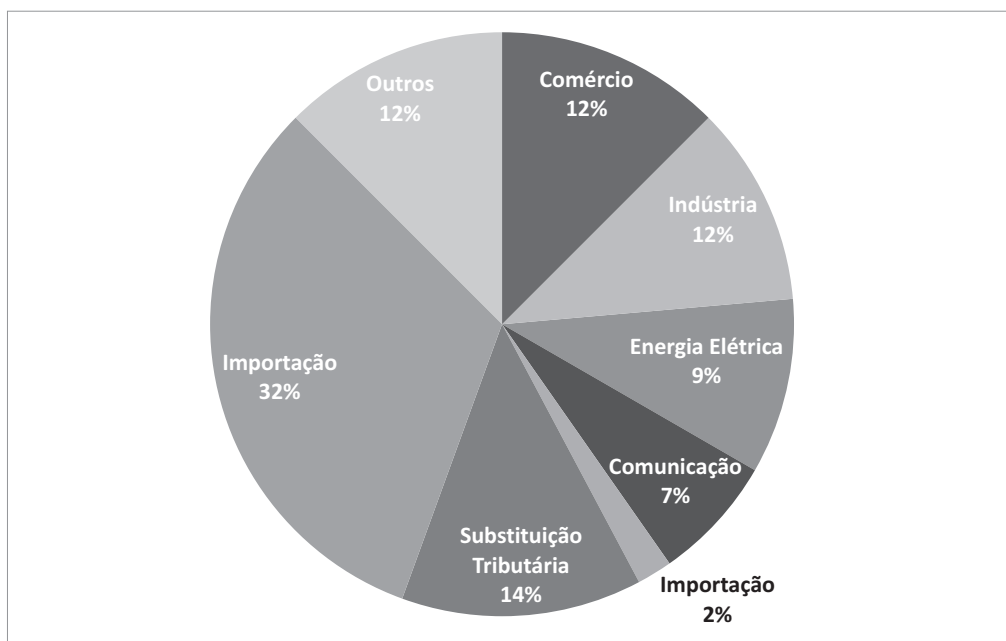
1. INTRODUÇÃO

O objetivo deste relatório é fornecer uma metodologia inicial para projetar as receitas de ICMS do Estado do Espírito Santo com base em modelos econométricos. Para tal, buscamos identificar as origens das receitas e definir variáveis antecedentes relacionadas a estas.

Teríamos duas abordagens gerais para nossos modelos de projeção, uma *top-down*, somente com o ICMS agregado e outra *bottom-up*, com projeções de receitas de ICMS setoriais. Dado a quantidade de desafios técnicos que serão expostos adiante e devido a incertezas quanto aos ganhos de abordagens mais desagregadas, optamos pela projeção das receitas agregadas de ICMS.

Ainda assim, o estudo da origem das receitas é importante para definição de variáveis que possam nos ajudar a projetar o agregado. O gráfico abaixo mostra que os setores com maiores contribuições, combinando mercadorias e serviços, são importações, comércio, indústria e substituição tributária; variáveis representativas destes setores podem ser de grande valia adiante.

Gráfico 1 - Origens das Receitas de ICMS (Média de 1999 a 2009)



Fonte: Governo do Estado do Espírito Santo.

Além dessa breve introdução, este relatório possui mais três seções. Na seção dois, à luz do discutido nesta introdução, é apresentado o modelo básico e são discutidas as metodologias de estimação. A seção três apresenta os resultados e define o melhor modelo para projeção, ao passo que a seção quatro conclui apresentando a projeção para a arrecadação pelo ICMS.

2. MODELO BÁSICO E METODOLOGIA DE ESTIMAÇÃO

Tomando por base o conteúdo da seção anterior, nesta seção iremos apresentar o modelo básico sobre o qual projetaremos as receitas de ICMS do Estado do Espírito Santo e definir a metodologia para essa projeção. Os setores de importação, comércio, indústria e substituição tributária são os que possuem maior contribuição para a arrecadação do ICMS, correspondendo a aproximadamente 70% das receitas no período amostral considerado. Um modelo que considere variáveis explicativas destes setores deveria ter, portanto, boa capacidade preditiva.

É interessante perceber que variáveis de importação, indústria e comércio do estado do Espírito Santo são facilmente obtidas e bem definidas, mas o mesmo não é válido para a substituição tributária. Assim sendo, optou-se por não utilizar variáveis deste setor no modelo de previsão¹. Todas as variáveis empregadas estão em logaritmos naturais e foram dessazonalizadas utilizando X-12 ARIMA², com a base de dados escolhida entre janeiro de 1999 e julho de 2009. O modelo básico tem a seguinte forma funcional genérica:

$$ICMS = \alpha_0 + \alpha_1 DET + \sum_{i=0}^3 \sum_{j=0}^1 \alpha_{i,j} X_{-j} + u_t \quad (1)$$

onde:

(i) *DET* são quaisquer variáveis determinísticas, como tendências lineares.

(ii) *X* é um vetor de variáveis explicativas, nominalmente importações (fonte: MDIC), produção industrial (fonte: PIM-PF, IBGE) e consumo (fonte: PMC, IBGE).

Dado o modelo básico, definimos agora a metodologia utilizada para a estimação das projeções, seguindo quatro etapas: testes de estacionariedade das variáveis envolvidas, estimações a partir de metodologias alternativas, inferência dos coeficientes obtidos e seleção do melhor modelo. Destaca-se a preocupação com problemas que possam aparecer em estimações com pequenas amostras, conforme é o caso do presente trabalho.

Em primeiro lugar, avalia-se a estacionariedade das variáveis escolhidas. Tal análise é de extrema importância na correta estimação das relações de equilíbrio entre as variáveis. Deve-se reconhecer a baixa potência dos testes de raiz unitária em amostras pequenas, o que pode levar a resultados equivocados que comprometam a estimação dessas relações de equilíbrio. Procurando minimizar tais riscos, definiu-se um procedimento em duas etapas, buscando aumentar a confiabilidade dos resultados.

¹ Ficará mais claro adiante que isso não implicará em penalização relevante da capacidade preditiva dos modelos (N.A.).

² Utilizamos o *software* DEMETRA, que permite definir corretamente os feriados brasileiros e nos fornece a especificação do X-12 ARIMA estimado, o que será de grande valia para a estimação da classe ARIMA de modelos de projeção.

A primeira etapa consistiu na estimação, para cada variável, dos testes PP (Phillips & Perron, 1988) e KPSS (Kwiatkowski, Phillips, Schmidt & Shin, 1992). Em ambos os testes, foram utilizados o núcleo espectral quadrático e janela de Andrews, comprovadamente mais eficientes em amostras pequenas (Andrews, 1991). A opção por tais testes foi baseada em seu bom comportamento para pequenas amostras e em suas hipóteses de teste alternativas. Enquanto o teste PP define como hipótese nula a presença de raiz unitária, o KPSS define estacionariedade. A estimação comparada destes forneceria, portanto, dois caminhos alternativos para avaliar a ordem de integração de uma determinada variável.

Na hipótese de resultados díspares, ou seja, quando cada um dos testes sugerir uma ordem de integração diferente para uma mesma variável, será estimado o teste padrão da literatura, o ADF (Dickey & Fuller, 1979), para definir a ordem de integração. A seguinte tabela mostra os resultados para os testes PP, KPSS e ADF, o último sempre que necessário, e a ordem de integração das variáveis. Todas as variáveis são estacionárias somente em primeira diferença.

Tabela 1 - Testes de Raíz Unitária

	PP	KPSS	ADF	INTEGRAÇÃO
ICMS	-0.954	0.762***		I (1)
Δ (ICMS)	-8.657***	0.103		I (0)
PMC	-1.603	0.159**		I (1)
Δ (PMC)	-3.339**	0.207		I (0)
PIM	-1.776	0.084	-0.866	I (1)
Δ (PIM)	-4.287***		-4.937***	I (0)
IMPORT	-1.517	0.462**		I (1)
Δ (IMPORT)	-4.744***	0.193		I (0)

*** Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Testes PP e KPSS estimados com núcleo espectral quadrático e janela de Andrews. Teste ADF estimado com número de lags (máximo de 30) escolhidos para minimizar o Critério de Informação de Scharwz. P-valores unicaudais obtidos de MacKinnon (1996).

Definidas as ordens de integração das variáveis, o passo seguinte foi a estimação propriamente dita dos modelos de projeção. Foram utilizados dois tipos distintos de metodologias, não necessariamente empregando todas as possíveis variáveis explicativas descritas na seção anterior. A primeira metodologia envolve estimações da classe ARIMA, que não utilizam qualquer tipo de estrutura, ou seja, são modelos estritamente com defasagens da variável dependente ICMS (componente AR) e dos resíduos estimados (componente MA).

A segunda metodologia de estimação avalia distintas combinações das variáveis explicativas descritas anteriormente. Partindo de uma regressão estática padrão por *Ordinary Least Squares* (OLS), há evidência de autocorrelação dos resíduos nessa regressão estática, o que pode nos levar a interpretações equivocadas para a relação de equilíbrio entre as variáveis escolhidas. Tal problema costuma ser comum em amostras pequenas e é sugerido pela própria estimação ARIMA já feita.

Foram escolhidos dois métodos para resolver esta questão. O primeiro foi a inclusão da variável ICMS defasada como variável explicativa das regressões OLS, o que em grande medida ajudou a resolver o problema de autocorrelação dos resíduos. O segundo foi a estimação de regressões dinâmicas para estabelecer a relação entre as variáveis. Dentre diversas opções, optou-se pelo *Dynamic Ordinary Least Squares* (DOLS) de Stock & Watson (1993), que nada mais é do que a estimação, por OLS, de um modelo aumentado pela primeira diferença das variáveis não estacionárias e de um dado número de *lags* e *leads* destas diferenças (denominado ordem), escolhido pela minimização do critério de informação de Schwarz, seguindo o modelo genérico abaixo:

$$Y = \alpha_0 + \alpha_1 Z + \alpha_2 W + \Delta W + \sum_{j=1}^p \alpha_j \Delta W_{-j} + \sum_{j=1}^q \alpha_j \Delta W_{+j} + u_t \quad (2)$$

onde:

$$Y \sim I(1)$$

$$Z \sim I(0)$$

$$W \sim I(1)$$

O próximo passo foi a inferência dos coeficientes obtidos nas regressões acima. Na presença de distúrbios esféricos, ou seja, homocedásticos e sem autocorrelação nos resíduos, ela é feita pela estatística *t* usual, em última instância, derivada de uma distribuição normal padrão. Há, portanto, dois problemas imediatos. Como já dito, há forte evidência de autocorrelação residual na estimação OLS usual e, além disso, a aproximação normal é tão melhor quanto maior é a amostra utilizada para a estimação, mas não se sabe *a priori* o tamanho de amostra mínimo para que a inferência seja adequada.

Quanto ao primeiro problema, além dos métodos de estimação utilizados para corrigir a autocorrelação dos resíduos descritos anteriormente, utilizou-se a matriz de variância-covariância de Newey-West, que fornece estimativas robustas para a estatística *t* e permite, portanto, inferência adequada, mesmo na presença de correlação residual.

Tratando das questões do tamanho da amostra, como já mencionado, nossa base de dados começa no início de 1999, o que nos dá pouco mais de 130 dados mensais ou pouco mais de 30 dados trimestrais. À primeira vista, as questões referentes ao tamanho da amostra estariam resolvidas com regressão em dados mensais; entretanto, estes estão sujeitos a uma variabilidade muito maior do que dados trimestrais, o que pode trazer ruído desnecessário para nossas estimativas. Mesmo reconhecendo suas possíveis restrições, julgou-se que os modelos trimestrais são suficientemente bem comportados e, portanto, trimestres serão nossa escolha primordial para frequência dos dados. Ainda assim, procurando uma melhor qualificação dos resultados obtidos, apresentaremos também os resultados para as estimações com frequência mensal.

A última etapa de nossa estratégia de estimação é definir critérios para a seleção dos melhores modelos, dentre todos os estimados. A literatura tem evoluído na direção do procedimento GETS, de Hendry & Krolzig (2001) e Hendry (2003b), baseado na Teoria de Redução do próprio Hendry³. De maneira geral este tem se mostrado melhor do que critérios tradicionais de escolha, como os de informação Akaike e Schwarz. Entretanto, exige uma amostra considerável para ser adequadamente implementado, o que foge ao escopo desse trabalho. Hendry & Krolzig (2001) ressaltam que a potência da seleção do melhor modelo cai com amostras pequenas, mas, mesmo assim, a minimização do critério de Schwarz ainda seria um bom procedimento para a escolha da melhor representação, independente da qualidade das hipóteses sobre as quais a especificação é construída.

Neste sentido, a escolha dos melhores modelos recai neste trabalho sobre a combinação de duas estatísticas de controle: a minimização do critério de informação de Schwarz e a maximização do R^2 ajustado, este último que mede o ajuste dos modelos estimados penalizando a inclusão de mais variáveis explicativas e, portanto, a diminuição dos graus de liberdade na estimação. Na próxima seção serão apresentados os resultados das estimações.

³ Para referências, vide Campos, Ericsson e Hendry (2003).

3. RESULTADOS

Conforme descrito na seção anterior, nossos modelos de previsão foram estimados utilizando três tipos de metodologias. Nossa estratégia nesta seção será apresentar todos os resultados e, posteriormente, definir qual é o melhor modelo de projeção.

3.1. Modelos AR(p)

Inicialmente, são apresentados os resultados da estimação dos modelos ARIMA para os modelos trimestrais. Nossos testes começaram com modelos que usaram estritamente a componente autorregressiva da variável ICMS. Dentre as opções avaliadas, observa-se na tabela 2 que o modelo trimestral com melhor ajuste é o AR(2), ou seja, que utiliza duas defasagens da variável dependente. A análise dos resíduos nesta especificação não sugere a presença de autocorrelação, logo não foi necessário utilizarmos a componente MA(q).

Tabela 2 - Modelos Trimestrais AR(p) – ICMS

Modelos AR(p)	C	AR(1)	AR(2)	AR(3)	AR(4)	R ² -adj	Schwarz
AR(1)	7.351***	0.906***				0.82	-1.79
AR(2)	7.428***	0.629***	0.252*			0.87	-2.17
AR(3)	7.472***	0.726***	0.129	0.059		0.86	-2.10
AR(4)	7.573***	0.684***	0.1922	-0.038	0.099	0.85	-2.00

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por OLS padrão, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

É interessante perceber que o coeficiente da primeira defasagem estimado no modelo AR(1) é bastante próximo de um, o que corrobora os resultados obtidos para os testes de raiz unitária descritos na seção anterior e levanta a hipótese de estimarmos o modelo em primeiras diferenças da variável dependente (DICMS). Novamente, a componente MA(q) não foi necessária.

Como se pode observar da tabela 3, o ajuste destes novos modelos em primeira diferença é bastante pobre, o que nos leva à escolha, dentro dos modelos trimestrais da classe ARIMA, da especificação ARIMA (2,0,0) já apresentada como a melhor para este tipo de metodologia de estimação.

Tabela 3 - Modelos Trimestrais AR(p) – D (ICMS)

Modelos AR(p)	C	AR(1)	AR(2)	AR(3)	AR(4)	R ² -adj	Schwarz
AR(1)	0.018**	-0.318**				0.109	-2.15
AR(2)	0.016*	-0.208	-0.090			0.001	-2.14
AR(3)	0.014*	-0.285*	-0.076	-0.121		0.006	-2.07
AR(4)	0.012	-0.364*	-0.207	0.002	-0.012	0.025	-2.07

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por OLS padrão, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

O resultado das estimações ARIMA para modelos mensais parecem, a princípio, um tanto distintos dos obtidos nos modelos trimestrais, possivelmente resultado da própria frequência dos dados. Mas, uma observação mais cuidadosa acaba por reforçar os resultados obtidos na estimação trimestral.

Tal qual anteriormente, nossos testes começaram por modelos AR(p) que utilizam somente defasagens da variável dependente ICMS. Observa-se na tabela 4 que o melhor modelo dessa classe foi aquele que utilizou até seis defasagens da variável ICMS, com resíduos nessa especificação que não denotam autocorrelação e, portanto, não fazem necessária a utilização de termos MA(q). Note que este resultado está alinhado ao resultado de frequência trimestral, com dois trimestres de defasagem como a melhor especificação.

Tabela 4 - Modelos Mensais AR(p) – ICMS

Modelos AR(p)	C	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)	R ² -adj	Schwarz
AR(1)	6.030***	0.879***						0.80	-1.16
AR(2)	6.119***	0.449***	0.475***					0.84	-1.38
AR(3)	6.245***	0.247***	0.289***	0.410***				0.86	-1.53
AR(4)	6.284***	-0.231**	0.280***	0.405***	0.031			0.86	-1.49
AR(5)	6.409***	0.221**	0.155*	0.317***	-0.046	0.303***		0.87	-1.54
AR(6)	6.340***	0.176**	0.171**	0.300***	-0.079	0.224***	0.134*	0.87	-1.65

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por OLS padrão, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

Tal qual nos modelos trimestrais, nos modelos mensais o coeficiente da primeira defasagem da variável dependente mostrou-se muito próximo de um, sugerindo a estimação de modelos em primeira diferença. Como se pode observar na tabela 5, ao contrário do ocorrido com os modelos trimestrais, o ajuste dos modelos mensais não foi tão pobre nessa especificação – mesmo que pior do que aquele com as variáveis em nível. O melhor modelo com resíduos sem evidência de autocorrelação é aquele que utiliza cinco defasagens da variável dependente.

Tabela 5 - Modelos Mensais AR(p) – D (ICMS)

Modelos AR(p)	C	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)	R ² -adj	Schwarz
AR(1)	0.007	-0,507***						0.25	-1.38
AR(2)	0.007	-0.729***	-0.421***					0.38	-1.55
AR(3)	0.007	-0.750***	-0.452***	-0.037				0.38	-1.51
AR(4)	0.007	-0.769***	-0.593***	-0.263**	-0.306***			0.44	-1.56
AR(5)	0.008***	-0.798***	-0.608***	-0.292**	-0.364***	-0.130		0.47	-1.64
AR(6)	0.007**	-0.736***	-0.529***	-0.235*	-0.304**	-0.049	0.105	0.43	-1.63

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por OLS padrão, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

Comparando o resultado dos modelos AR(p) mensais tanto em nível como na primeira diferença da variável dependente, observa-se que o melhor modelo é o da classe ARIMA (6,0,0), ou seja, com a variável em nível e com seis defasagens explicativas. Portanto, temos como melhores especificações duas defasagens nos modelos trimestrais e seis defasagens nos modelos mensais, sempre com as especificações com variáveis em nível dominando aquelas em primeira diferença.

3.2. Modelos OLS com Defasagem da Variável Dependente

Definidos os melhores modelos ARIMA, agora nos voltamos para os modelos que buscam explicar o comportamento do ICMS através das variáveis explicativas descritas na seção 2 (indústria, comércio e importações). Como já exposto, utilizamos duas abordagens para estimar a relação entre essas variáveis, sempre buscando resolver o problema de autocorrelação nos resíduos das regressões.

Os resultados para a família de modelos com estimação trimestral por OLS padrão e uma defasagem da variável dependente (portanto, uma evolução do modelo AR(1)) são resumidos na tabela 6, onde apresentamos algumas das combinações possíveis, de acordo com a defasagem utilizada nas variáveis explicativas⁴.

⁴ Por restrições amostrais, resultado da frequência de dados escolhida, nos limitamos ao máximo de uma defasagem para cada variável. A variável determinística de tendência linear não foi significativa em nenhum modelo estimado e, portanto, foi retirada da versão final apresentada. Os modelos mensais, apresentados adiante, terão o número máximo de defasagens limitado a três, a fim de manter consistência com a máxima defasagem (um trimestre) dos modelos trimestrais.

Há um ponto que merece destaque. O estudo da origem das receitas de ICMS do Estado do Espírito Santo (rapidamente discutido na introdução deste trabalho) sugere a proeminência das importações como variável explicativa, na medida em que correspondia a praticamente 1/3 das receitas. Entretanto, em nenhuma especificação as importações se mostraram significantes⁵. A justificativa para tal padrão inesperado estaria na capacidade que as variáveis de indústria e consumo teriam em descrever diretamente a dinâmica da atividade que está por trás das próprias importações.

Atendo-se à interação das defasagens de indústria e comércio, se observa que a melhor especificação, de acordo com os critérios de seleção definidos na seção anterior, é aquela onde ambas as variáveis estão defasadas por um trimestre (destacado em *itálico e negrito* na tabela 6). As elasticidades obtidas são altamente significantes, principalmente a da indústria, e com sinal positivo, conforme esperado. Além disso, a componente inercial, medida pela variável dependente defasada também é altamente significativa, com o sinal intuitivo. Por fim, apesar de não utilizar a variável de importações, o modelo possui bom ajuste aos dados, com R^2 ajustado maior que 90%⁶.

⁵ Isso ocorreu tanto quando importações eram a única variável explicativa como quando eram acompanhadas por outras variáveis. Assim sendo, optamos por apresentar no estudo somente modelos onde importações aparecem com outras variáveis. Outras estimativas podem ser obtidas junto aos autores.

⁶ Uma vantagem da especificação vencedora é que as projeções podem ser feitas sem a necessidade de definir cenários para as variáveis explicativas, posto que os valores utilizados são defasados; assim, a projeção um trimestre adiante das receitas do ICMS precisa dos valores da receita de ICMS, da produção industrial e do comércio do Estado contemporâneos, dados já conhecidos no momento de rodar as projeções. Voltaremos a este ponto adiante.

Tabela 6 - Modelos Trimestrais com Variáveis Explicativas e Defasagem da Variável Dependente (OLS)

Defasagens (PIM, PMC, IMPORT)	C	ICMS (-1)	PIM (.)	PMC (.)	IMPORT (.)	R ² -adj	Schwarz
	0.387	0.693***	0.392***			0.85	-1.91
(0,0,0)	1.082**	0.471***	0.191**	0.378***		0.88	-2.34
	1.093**	0.482***	0.197**	0.351***	0.007	0.88	-2.24
(0,-1,0)	0.653*	0.563***	0.257***	0.274**		0.89	-2.37
	0.675*	0.564***	0.259***	0.257**	0.004	0.89	-2.27
(0,0,-1)	1.075**	0.482***	0.201**	0.354***	0.005	0.88	-2.24
(0,-1,-1)	0.641*	0.559***	0.248***	0.296***	-0.006	0.89	-2.27
	0.547	0.563***	0.559***			0.88	-2.08
(-1,0,0)	1.087**	0.386***	0.370***	0.342***		0.90	-2.54
	1.106**	0.389***	0.374***	0.316***	0.007	0.90	-2.45
(-1,-1,0)	0.728*	0.470***	0,411***	0.246**		0.91	-2.58
	0.742*	0.471***	0,412***	0.235**	0.003	0.91	-2.49
(-1,0,-1)	1.097**	0.393***	0.389***	0.288**	0.012	0.90	-2.45
(-1,-1,-1)	0.724*	0.470***	0.410***	0.249**	0.000	0.90	-2.49

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por OLS padrão, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

A estimação dos modelos OLS mensais possui duas grandes diferenças em relação àquela do modelo trimestral. Em primeiro lugar, falando dos coeficientes, as importações mostraram-se significantes na maior parte das especificações testadas, mas invariavelmente com sinal inesperado (negativo), mesmo que com coeficiente baixo⁷. Além disso, a inclusão da variável determinística de tendência, em geral, piorou as regressões e, por causa disso, acabamos por retirá-la da especificação final.

Em segundo lugar, ressaltamos a relação existente entre a defasagem da variável dependente necessária para estabilizar os resíduos dos modelos mensais e trimestrais. Testes no modelo cheio (ou seja, com todas as variáveis explicativas: indústria, comércio e importações) indicaram que a defasagem da variável dependente adequada para estabilizar os resíduos da regressão foi de três períodos, o que, de certa forma, está alinhado à defasagem indicada nos modelos trimestrais, que era de um trimestre. As regressões mensais foram estimadas, portanto, com uma defasagem de três meses na variável ICMS e com defasagens de até três meses das variáveis explicativas.

A tabela 7 apresenta resumidamente os resultados destas regressões, demonstrando somente os resultados de combinações possíveis das variáveis explicativas defasadas em três meses⁸. De acordo com os critérios de seleção definidos na seção anterior, o melhor modelo mensal é aquele que utiliza a defasagem de três meses das variáveis de indústria e comércio, sem importações e tendência determinística (destacado em negrito na tabela 7). Tal qual nos modelos trimestrais, as elasticidades obtidas foram altamente significantes, com o sinal positivo esperado, o mesmo ocorrendo com a elasticidade da componente inercial. O ajuste do modelo também parece elevado, novamente com R^2 ajustado ao redor de 90%.

Observa-se que, tal qual ocorrido nos modelos $AR(p)$, há uma grande relação entre os resultados das estimações mensais e trimestrais. De fato, ambas as periodicidades amostrais levaram à mesma especificação escolhida, com valores defasados das variáveis explicativas comércio e indústria. As elasticidades estimadas para a indústria foram praticamente as mesmas, ao passo que a da componente inercial foi razoavelmente próxima. A maior diferença fica por conta da participação do comércio, com elasticidade mensal que é praticamente o dobro da trimestral.

⁷ Como veremos adiante, tal fato acaba por não ser um problema, na medida em que a melhor especificação, definida de acordo com os critérios de informação que já discutimos, não possui a variável de importações.

⁸ Os modelos foram estimados para todas as combinações de defasagens das variáveis explicativas, mas para facilitar a apresentação só foram demonstradas combinações com a defasagem do modelo vencedor. Todos os outros resultados podem ser obtidos juntos aos autores (N.A.).

Tabela 7 - Modelos Mensais com Variáveis Explicativas e Defasagem da Variável Dependente (OLS)

Defasagens (PIM, PMC, IMPORT)	C	ICMS (-3)	PIM (.)	PMC (.)	IMPORT (.)	R ² -adj	Schwarz
(0,0,0)	-0.184	0.708***	0.415***			0.84	-1.42
	-0.474*	0.413***	0.201***	0.634***		0.88	-1.83
	-0.588**	0.406***	0.187**	0.700***	-0.02**	0.88	-1.81
(0,-3,0)	-0.474*	0.439***	0.236***	0.569***		0.88	-1.85
	-0.542*	0.437***	0.230***	0.604***	-0.010	0.88	-1.82
(0,0,-3)	-0.574**	0.383***	0.156*	0.762***	-0.026**	0.88	-1.82
(0,-3,-3)	-0.588**	0.417***	0.213***	0.663***	-0.020*	0.88	-1.84
(-3,0,0)	-0.401	0.609***	0.590***			0.86	-1.53
	-0.706***	0.359***	0.377***	0.578***		0.89	-1.93
	-0.844***	0.352***	0.371***	0.642***	-0.019***	0.89	-1.92
(-3,-3,0)	-0.655***	0.392***	0,388***	0.517***		0.89	-1.95
	-0.749***	0.387***	0,387***	0.557***	-0.012	0.89	-1.92
(-3,0,-3)	-0.836***	0.331***	0.349***	0.691***	-0.024**	0.89	-1.93
(-3,-3,-3)	-0.788***	0.369***	0.373***	0.609***	-0.021*	0.89	-1.94

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por OLS padrão, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

3.3. Modelos DOLS

Dando sequência aos métodos utilizados para resolver o problema de autocorrelação residual das regressões, apresentam-se agora os resultados de estimação dos modelos com variáveis explicativas defasadas através de DOLS. A ideia por trás da estimação dinâmica é permitir a estabilização dos resíduos das regressões sem a utilização da variável dependente defasada, fazendo com que a variável dependente seja avaliada somente por sua interação com as variáveis explicativas escolhidas.

Os resultados para a família de modelos com periodicidade trimestral são resumidos na tabela 8, onde novamente apresentamos diversas combinações de defasagens das variáveis explicativas⁹.

Nesta metodologia, de acordo com os critérios de seleção de modelos definidos, a melhor especificação é aquela que utiliza os valores de comércio contemporâneos e os da indústria defasados em um trimestre (destacado em *itálico e negrito* na tabela 8). As elasticidades obtidas possuem os sinais esperados e são altamente significantes, com elevado ajuste do modelo aos dados (R^2 ajustado acima de 90%). Tal qual nos modelos OLS trimestrais, em nenhuma especificação testada as importações parecem relevantes; a justificativa para tal padrão apresentada anteriormente ainda se aplica.

Procurando manter a consistência com as defasagens propostas nos modelos trimestrais, os modelos DOLS mensais foram estimados com defasagem de até três meses nas variáveis explicativas. Os resultados obtidos foram similares aos observados nos modelos trimestrais.

Em primeiro lugar, a variável de importação não foi significativa na imensa maioria das especificações e, quando o foi, teve nível de significância baixo e coeficiente próximo de zero. Além disso, as elasticidades obtidas para comércio e indústria tiveram os sinais esperados, mesmo que, em geral, com valores distintos das equivalentes nos modelos trimestrais. Segundo os critérios de escolha já definidos, a melhor especificação foi a que utilizou defasagem de três meses na variável de indústria e de um mês na variável de comércio, como se pode observar na tabela 9¹⁰. Note que foi mantida a diferença de defasagens entre comércio e indústria observada nos modelos trimestrais.

Apresentamos, portanto, os melhores modelos estimados de acordo com cada metodologia e defasagem. A próxima seção dedica-se a definir um critério de seleção entre essas metodologias e defasagens, indicando o modelo que será utilizado para nossas projeções.

⁹ Para facilitar a notação não serão reportados os coeficientes dos termos com *lags* e *leads* das variáveis explicativas. De fato, estes termos funcionam como um instrumento estatístico destinado a estabilizar os resíduos da regressão e, portanto, não teriam intuição econômica alguma; sendo assim, não serão utilizados em nossos modelos de projeção.

¹⁰ Dada a pouca relevância das importações, optamos por não apresentá-la na tabela. Assim sendo, a tabela 9 mostra os resultados para distintas defasagens das variáveis de indústria e comércio. Novamente, os coeficientes dos termos defasados não foram apresentados, por razões já expostas. A lista completa de modelos estimados pode ser obtida junto aos autores.

Tabela 8 - Modelos Trimestrais com Variáveis Explicativas (DOLS)

Defasagens (PIM, PMC, IMPORT)	C	PIM (.)	PMC (.)	IMPORT (.)	R ² -adj	Schwarz
	1.872***	1.147***			0.86	-2.00
(0,0,0)	2.046***	0.582**	0.509***		0.91	-2.21
	2.370***	0.585*	0.407	0.026	0.90	-1.88
(0,-1,0)	2.055***	0.717***	0.376***		0.91	-2.24
	3.054***	0.799**	-0.058	0.114	0.92	-2.12
(0,0,-1)	2.046***	0.598	0.489	0.004	0.88	-1.85
(0,-1,-1)	2.454***	0.826**	0.095	0.070	0.91	-2.05
	1.849***	1.154***			0.85	-1.94
(-1,0,0)	2.163***	0.522***	0.543***		0.91	-2.26
	2.498***	0.481**	0.491**	0.018	0.91	-1.96
(-1,-1,0)	1.948***	0.670***	0,445***		0.90	-2.24
	3.066***	0.675***	0,094	0.087	0.92	-2.14
(-1,0,-1)	2.150***	0.515**	0.553*	-0.001	0.90	-1.92
(-1,-1,-1)	2.122***	0.724**	0.305	0.039	0.90	-1.88

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por DOLS de ordem 1, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

Tabela 9 - Modelos Mensais com Variáveis Explicativas (DOLS)

Defasagens (PIM, PMC)	C	PIM (.)	PMC (.)	R ² -adj	Schwarz
(0,_)	-1.020**	1.497***		0.70	-0.71
(0,0)	-0.793**	0.366***	1.053***	0.86	-1.53
(0,-1)	-0.694**	0.348***	1.052***	0.87	-1.61
(0,-2)	-0.735*	0.379***	1.030***	0.86	-1.56
(0,-3)	-0.764**	0.398***	1.018***	0.85	-1.51
(-1,_)	-1.074***	1.509***		0.73	-0.81
(-1,0)	-0.908***	0.455***	0.990***	0.87	-1.58
(-1,-1)	-0.789**	0.417***	1.003***	0.87	-1.65
(-1,-2)	-0.843**	0.454***	0.979***	0.87	-1.59
(-1,-3)	-0.869**	0.472***	0.968***	0.86	-1.56
(-2,_)	-1.108***	1.517***		0.74	-0.88
(-2,0)	-1.000***	0.513***	0.953***	0.87	-1.58
(-2,-1)	-0.886***	0.478***	0.964***	0.87	-1.65
(-2,-2)	-0.916**	0.505***	0.945***	0.87	-1.60
(-2,-3)	-0.947**	0.526***	0.931***	0.86	-1.57
(-3,_)	-1.111***	1.518***		0.75	-0.94
(-3,0)	-1.033***	0.551***	0.923***	0.87	-1.58
(-3,-1)	-0.921***	0.509***	0.942***	0.87	-1.66
(-3,-2)	-0.958***	0.546***	0.914***	0.87	-1.62
(-3,-3)	-0.966***	0.559***	0.903***	0.86	-1.58

***Rejeição a 1%.

** Rejeição a 5%.

* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por DOLS de ordem até 3, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

3.4. Escolha do Modelo de Projeção

Antes de definirmos nosso modelo de projeção, cabe uma consideração importante. Os modelos vencedores para cada metodologia (ARIMA, OLS e DOLS) e frequência (trimestral e mensal) empregadas possuem termos defasados e/ou contemporâneos aos dados de arrecadação pelo ICMS, o que traz alguns desafios para as projeções.

Modelos com variáveis defasadas têm a vantagem de possuir toda a base de dados relevante conhecida no tempo T , o que permite a projeção imediata em $T+1$ utilizando os dados correntes. Entretanto, esses modelos não são particularmente bons para prever momentos de reversão de tendência, posto que, por construção, possuem uma inércia muito grande. Modelos com variáveis contemporâneas não teriam esse problema, mas a projeção adiante necessita de hipóteses sobre as variáveis explicativas em $T+1$.

Ressaltadas estas questões, a tabela 10 compila os melhores modelos para cada metodologia e compara os resultados obtidos para as frequências mensal e trimestral, além das estatísticas de informação relacionadas. É possível entender os modelos estimados por OLS derivando dos modelos ARIMA, incorporando variáveis explicativas independentes. A estimação por DOLS seria uma evolução do modelo estimado por OLS, permitindo avaliar somente o papel das variáveis explicativas, sem a necessidade de utilizar valores defasados do ICMS para estabilizar os resíduos das regressões.

Primeiramente nota-se que os modelos trimestrais dominam estritamente os modelos mensais, em geral apresentando tanto R^2 ajustado quanto estatística de informação melhores. Como já abordamos anteriormente, a escolha entre frequências mensais e trimestrais pressupunha uma escolha explícita entre uma maior amostra para estimação e maior estabilidade amostral. Os resultados sugerem que, ao menos nesse caso, os ganhos com a estabilidade foram maiores do que os da ampliação da amostra.

Definida a melhor frequência de dados, passamos à escolha do melhor modelo. Pelos critérios de seleção definidos, imediatamente podemos descartar o modelo ARIMA trimestral, que possui tanto o pior R^2 ajustado quanto o mais elevado critério de informação de Schwarz. Os modelos OLS e DOLS trimestrais são equivalentes em termos de ajuste do modelo aos dados e a variável comum aos dois (PIM defasada em um trimestre) possui elasticidades muito similares. Entretanto, observa-se que, pelo critério de informação de Schwarz, a especificação estimada por OLS parece melhor do que aquela estimada por DOLS.

Assim, o modelo OLS(-1,-1,-1), destacado em itálico e negrito na tabela 10, será nosso modelo de projeção para a arrecadação de ICMS do estado do Espírito Santo.

Tabela 10 - Comparação de Modelos e Frequência de Dados (ARIMA, OLS, DOLS)

Modelos	Variáveis							R ² -adj	Schwarz
	C	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)		
ARIMA									
Mensal (-6,0,0)	6.340***	0.176**	0.171**	0.300***	-0.079	0.224***	0.134***	0.87	-1.65
Trimestral (-2,0,0)	7.428***	0.629***	0.252*					0.87	-2.17
OLS	C	ICMS(.)		PIM(.)		PMC(.)		R²-adj	Schwarz
Mensal (-3,-3,-3)	-0.655***	0.392***		0.388***		0.517***		0.89	-1.95
Trimestral (-1,-1,-1)	0.728*	0.470***		0.411***		0.246**		0.91	-2.58
DOLS	C	ICMS(.)		PIM(.)		PMC(.)		R²-adj	Schwarz
Mensal (0,-3,-1)	-0.921***			0.509***		0.942***		0.87	-1.66
Trimestral (0,-1,0)	2.163***			0.522***		0.543***		0.91	-2.26

***Rejeição a 1%.

** Rejeição a 5%.

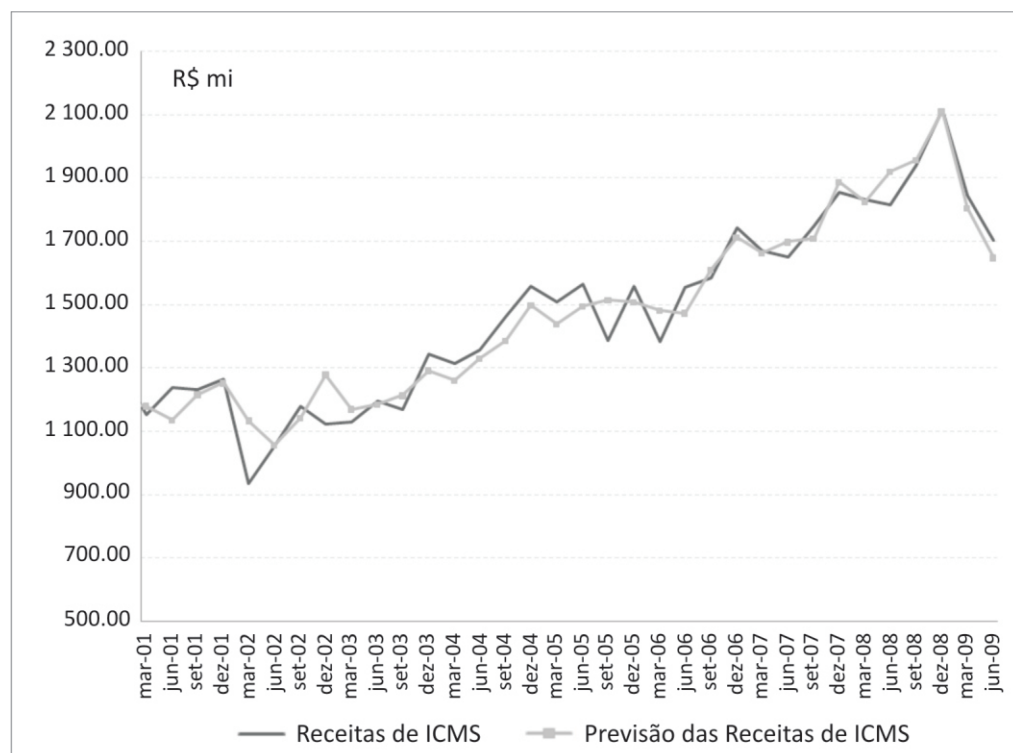
* Rejeição a 10%.

Variáveis em logaritmo natural dessazonalizadas por X-12 ARIMA. Estimação por OLS padrão ou DOLS de ordem até 3, com inferência utilizando a matriz de variância-covariância robusta de Newey-West. R²-adj é o coeficiente explicativo ponderado pelos graus de liberdade do modelo.

4. PROJEÇÃO E CONCLUSÃO

De acordo com o exposto na seção anterior, o melhor modelo dentre todos os apresentados é aquele, estimado por OLS, que define a receita de ICMS corrente como função da receita de ICMS defasada, da produção industrial defasada e do comércio defasado, todos por um trimestre. O ajuste do modelo aos dados originais pode ser observado no gráfico 2, com dados desde o início de 2001 até o segundo semestre de 2009.

Gráfico 2 - Previsão In-Sample



Fonte: Governo do Estado do Espírito Santo.

Elaboração: Cálculo dos Autores.

Para avaliar a capacidade preditiva do modelo fora da amostra, estimamos as previsões um passo adiante desde o primeiro trimestre de 2008. Na tabela 11, comparamos os resultados dessas projeções com os dados efetivamente observados e calculamos o erro percentual obtido. É interessante perceber que, em 2009, o modelo subestimou as receitas de ICMS.

Tabela 11 - Projeção Um Trimestre Adiante

	2008.1	2008.2	2008.3	2008.4	2009.1	2009.2	2009.3
Dado Efetivo	1832.87	1815.52	1938.74	2114.74	1846.06	1701.69	
Projeção	1824.78	1921.70	1958.18	2112.42	1804.87	1648.56	1663.13
Erro (%)	-0.44%	5.85%	1.00%	-0.11%	-2.23%	-3.12%	

Projeção por OLS, em milhões de reais. Níveis recuperados utilizando dummies sazonais sugeridas pelo modelo X-12 ARIMA.

Por fim, na mesma tabela apresentamos a previsão de arrecadação de ICMS para o terceiro trimestre do ano corrente. Nossa projeção é de R\$ 1663,13 milhões, o que seria consistente com uma queda de pouco mais de 14% em 12 meses, acima da queda ao redor de 8% observada no segundo trimestre. Entre o segundo trimestre efetivo e o terceiro trimestre estimado, a queda projetada com dados dessazonalizados seria ao redor de 4%.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ANDREWS, D.W.K (1991): Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation. *Econometrica*, V. 59, n. 3, p. 817-858.

CAMPOS, J.; ERICSSON, N.R. & HENDRY, D.F. (2003): Consistent Model Selection by an Automatic Gets Approach. *Oxford Bulletin of Economics and Statistics*, n.65, p. 803-819.

DICKEY, D. & FULLER, W.A. (1979): Distribution of the Estimates for Autoregressive Time Series with a Unit Root. *Journal of the American Statistical Association*, v.74, p.427-431.

HENDRY, D.F. (2003a): *The Properties of Automatic GETS Modelling*. Mimeo, Economics Department, Oxford University.

HENDRY, D.F. (2003b): New Developments in Automatic General-to-specific Modelling. In: STIGUM, B.P. (Ed.): *Econometrics and the Philosophy of Economics*. Princeton University Press.

HENDRY, D.F. & KROLZIG, H.M. (2001): Computer Automation of General-to-Specific Model Selection Procedures. *Journal of Economic Dynamics and Control*, n. 25, p. 831-866.

KWIATKOWSKI, D.; PHILLIPS, P.; SCHMIDT, P. & SHIN, Y. (1992): Testing the Null Hypothesis of Stionarity Against the Alternative of a Unit Root: How Sure are We that Economic Time Series have a Unit Root? *Journal of Econometrics*, v.54, p.159-178.

MACKINNON J. G. "Numerical distribution for unit root and cointegration tests." *Journal of Applied Econometrics*, 11 (1996), 601-618.

PHILLIPS, P. & PERRON, P. (1988): Testing for a Unit Root in Time Series Regression. *Biometrika*, v.75, p.335-346.

STOCK, J.H. & WATSON, M.W. (1993): A Simple Estimator of Cointegration Vectors in Higher Order Integrated Systems. *Econometrica*, v.61, p. 783-820.

Editoração
João Vitor André

Bibliotecária
Andreza Ferreira Tovar

Instituto
Jones dos Santos Neves - IJSN

Secretaria
de Economia
e Planejamento

UM NOVO

ESPÍRITO SANTO

Governo do Estado

www.es.gov.br